

# Rethinking Microblogging: Open, Distributed, Semantic

Alexandre Passant<sup>1</sup>, John G. Breslin<sup>1,2</sup>, and Stefan Decker<sup>1</sup>

<sup>1</sup> Digital Enterprise Research Institute, National University of Ireland, Galway  
`firstname.lastname@deri.org`

<sup>2</sup> School of Engineering and Informatics, National University of Ireland, Galway  
`john.breslin@nuigalway.ie`

**Abstract.** In order to break down the walls that lock-in social data and social networks, new paradigms and architectures must be envisioned. There needs to be a focus on the one hand on distributed architectures — so that users remain owners of their data — and on the other hand on means to semantically-enhance their content — so that it becomes more meaningful and interoperable. In this paper, we detail the anatomy of SMOB, a distributed semantic microblogging framework. In particular, we describe how it achieves the previous objectives using Semantic Web standards (including RDF(S)/OWL, RDFa, SPARQL) and Linked Data principles, as a consequence rethinking the microblogging experience and, more generally, providing Linked Social Data as part of the growing Linking Open Data cloud.

**Keywords:** Social Web, Semantic Web, Linked Data, Microblogging, Distributed Systems.

## 1 Introduction

Founded in 2006, Twitter<sup>1</sup> defined the foundations of a now well-known phenomena: microblogging. While blogs let people openly share their thoughts on the Web, microblogging goes further by enabling real-time status notifications and micro-conversations in online communities. While it is mainly recognised as a way to provide streams of information on the Web, it can be used in various settings such as Enterprise 2.0 environments. The simplicity of publishing microblogging updates, generally shorter than 140 characters, combined with the ubiquitous nature of microblog clients, makes microblogging an unforeseen communication method that can be seen as a hybrid of blogging, instant messaging and status notification. Moreover, by considering microblog content as being information streams, new real-time applications can be imagined in the realm of citizen sensing [18].

So far, most of the current research around microblogging focuses on studying and understanding its communication patterns [10] [11]. However, its technical

---

<sup>1</sup> <http://twitter.com>

and engineering aspects still have to be studied, especially regarding the need for more openness in Social Web applications, argued by various manifestos such as “A Bill of Rights for Users of the Social Web”<sup>2</sup>. This would enable means to make microblogging integrated more closely with other Social Web applications, as well as with other data from the Web, for instance to identify information about a particular topic or event. In this paper, we discuss the motivations, the architecture and the use of SMOB — <http://smob.me> —, an open-source framework for open, distributed, and semantic microblogging<sup>3</sup>. In particular, our contributions include:

- an ontology stack to represent microblogs and microblog posts, using and extending popular vocabularies, hence making microblog data integrated with other efforts from the Semantic Web realm;
- a distributed architecture, based on hubs interacting together using SPARQL and SPARQL/Update, and which could also be used in a broader context of real-time RDF data replication; and
- interlinking components, enhancing existing practices such as #tags and providing the means to make microblogging interlinked with any resource from the growing Linking Open Data (LOD) cloud<sup>4</sup>.

The rest of the paper is organised as follows. In Section 2, we discuss different engineering issues of microblogging systems, from which we derive three main requirements to enhance the microblogging experience. Based on these requirements, Section 3 describes the anatomy of SMOB, focusing on our main contributions and their applicability to other components of the Social Semantic Web. We then describe the system in use (Section 4), demonstrating how the previous principles enhance the microblogging experience in terms of content publishing and data reuse and discovery. We discuss related work in Section 5, and finally conclude the paper.

## 2 Issues with Current Microblogging Services and Requirements for a Richer User Experience

### 2.1 Lack of Machine-Readable Metadata

A first issue regarding microblogging services concerns the lack of metadata exposed from these applications, in a way that could be easily reused. Twitter has adopted microformats<sup>5</sup> for describing lists of followers (and followees), but this does not provide the means to match one username with other profiles that she or he may have in other platforms. Relatedly, users must create a new profile

<sup>2</sup> <http://opensocialweb.org/2007/09/05/bill-of-rights/>

<sup>3</sup> While the first release of SMOB has been designed mid-2008, this paper describes its recent 2.x series, redesigned with a completely new architecture and introducing new paradigms for semantic microblogging.

<sup>4</sup> <http://richard.cyganiak.de/2007/10/lod/>

<sup>5</sup> <http://microformats.org/>

— and fill in related details — on each platform they join. While OpenID could be used as a first step to solve this, there are still some issues with regards to maintaining attributes (profile pictures, etc.) across different websites. Furthermore, posts themselves do not provide enough fine-grained metadata (creation date, topics being addressed, recipient(s), etc.), beyond the use of RSS elements. Such a lack of metadata makes it difficult for microblogging to be interoperable with other systems, and it is also difficult for the data to be efficiently queried (*e.g.* to retrieve all microblog posts written by one’s friend in the last six days). Consequently, we identified a first requirement to enhance microblogging systems:

- *R1: machine-readable metadata* — in order to make microblogging more interoperable, not only between microblogging applications but also considering the Social Web at large, there is a need for more machine-readable metadata about (1) microblog posts (time, author, etc.), (2) their content (topics, etc.) and (3) their authors (name, depiction, etc.).

## 2.2 Microblogs as Closed-World Data Silos

A second issue is that microblogging services act as closed worlds similar to most Web 2.0 applications. To that extent, the current centralised architecture of microblogging systems raises an important issue in terms of data portability, since data is locked into a particular platform, and cannot be automatically reused in other applications. While subscribing to RSS feeds in order to get a local copy of one’s data can be an alternative, it does not provide the means to gather a complete dump of, *e.g.* last year’s Twitter activity in a particular domain, unless one crawled the feed since the beginning. Therefore, our second requirement is the following:

- *R2: decentralised architecture and open data* — in order to solve the walled-garden issue of current microblogging systems, new decentralised architectures must be provided, so that everyone can setup her or his own platform and claim ownership on the published data, while at the same time making it openly available for other applications that may require it.

## 2.3 Lack of Semantics in Microblog Posts

Finally, in addition to the previous metadata concerns, microblog posts themselves do not carry any semantics, making their querying and reuse difficult. Twitter users have adopted hashtags (**#tag** patterns included in microblog posts to emphasise particular words, now officially supported by the service, as is the **@user** pattern), but their semantics are not readily machine-processable, thus raising the same ambiguity and heterogeneity problems that tagging practices cause [14]. Someone interested in the Semantic Web would have to follow various tags to get an accurate description of what is happening in that realm. Especially, she or he will have to consider the different tag variations caused

by the sparse expertise levels and backgrounds of users, as raised by [7] in Delicious, *i.e.* subscribing to #RDF, #SPARQL, etc. This consequently leads to our third requirement:

- *R3: data-reuse and interlinking* — in order to enhance semantic descriptions of microblog posts and their content, they must be interlinked with existing resources using meaningful and typed relationships, going further than the traditional usage of ambiguous and heterogeneous hashtags and hyperlinks.

### 3 Anatomy of SMOB

Based on the aforementioned requirements, we engineered SMOB — Semantic MicroBlogging —, an open-source microblogging framework providing an open, distributed and semantic microblogging experience. In particular, the system completely relied on state-of-the-art Semantic Web and Linked Data [3] technologies to achieve this goal. Therefore, SMOB offers a combination of these technologies with existing Social Web paradigms and tools (microblogging, information streams), leading to what is generally known as the Social Semantic Web [5]. To enable such integration and solve the aforementioned issues, we have engineered a system relying on several components. In particular:

- *R1: machine-readable metadata* is achieved thanks to lightweight *ontologies* and RDFa markup, enabling common semantics and standard representations to model microblog posts (and microblog services) and their metadata and consequently providing interoperable descriptions of microblog posts. Thus, posts can be exchanged not only between SMOB applications but with any service capable of consuming RDF(S)/OWL data;
- *R2: decentralised architecture and open data* is achieved thanks to *distributed hubs*, spread across the Web and exchanging information (posts and subscriptions) based on the previous ontologies and a sync protocol (based on SPARQL/Update over HTTP). These hubs also act as end-users publishing and browsing interfaces for microblog posts;
- *R3: data-reuse and interlinking* is achieved thanks to *interlinking components*, so that microblog posts can be interlinked with resources from the Web, and in particular those from the aforementioned Linking Open Data cloud, by turning #tags into URIs identifying Semantic Web resources. Thus, it allows (1) microblog posts to become more discoverable, by being linked to existing resources using the Linked Data principles [2]; and (2) microblogging and the Social Web to join the Linking Open Data Cloud, and not exist as an isolated subset.

In addition, thanks to these different components, the boundaries between microblogging and other Social Semantic Web applications become (voluntarily) weaker. Microblog content can indeed be immediately mashed-up and integrated between various applications supporting the same standards. This emphasises the object-centric [13] nature of many online conversations and social networks,

enhancing them with a real-time component. For example, it provide means to integrate microblog updates with blog posts or RSS feeds mentioning the same object (research topic, project, etc.), this object being identified by its own URI.

Technically, as our goal was to make the system as easy as possible to deploy, SMOB only requires a LAMP — Linux, Apache, PHP and MySQL — environment. This also emphasises how Semantic Web and Linked Data technologies can be brought to end users and developers thanks to simple frameworks and engineering practices linking these technologies to object-oriented paradigms. In our case, we relied on the ARC2 PHP framework<sup>6</sup>, which provides a simple toolkit to build RDF-based applications in PHP. The different objects (posts, users, etc.) and views (lists of posts, map views, etc.) have been mapped to SPARQL and SPARQL/Update queries using the aforementioned ontologies in order to be generated and saved in the local database, so that the underlying RDF(S)/OWL structure is directly mapped to PHP objects.

### 3.1 The SMOB Ontologies Stack

In order to semantically-enhance microblogging services and microblog posts by providing more fine-grained metadata (R1), there is a need for models representing these services and their related content. This entails the need for ontologies representing:

- users, their properties (name, homepage, etc.), their social networking acquaintances and contextual information about themselves (geographical context, presence status, etc.);
- microblogging services and microblog posts, including common features such as `#tags` and replies (`@user`).

In order to model user profiles, we naturally relied on FOAF — Friend of a Friend [6] — as it provides a simple way to define people, their main attributes and their social acquaintances. Moreover, FOAF is already widely used on the Web, providing SMOB users with a way to reuse their existing profiles in their microblogs, thus also enabling distributed profile management and authentication. As we mentioned earlier, one major issue with current Web 2.0 services is the need to create a new profile each time one wants to join a new website. Given the popularity and the number of new services appearing regularly on the Web, this quickly becomes cumbersome and leads to what some have called “social network fatigue”. We have addressed this issue by letting users create a SMOB account simply by specifying their FOAF URI. User profiles are then retrieved from these URIs (assuming they follow the Linked Data principles), so that name of an author, depiction, etc. can be stored and updated in third-party websites, but are however integrated in microblog updates. In addition, the original profile can be any any RDF serialisation, for example in the form of an RDFa-annotated profile (Fig. 1). Furthermore, each post is linked to this author (using `foaf:maker`), every author being uniquely identified on the Web

---

<sup>6</sup> <http://arc.semsol.org>



Fig. 1. Distributed and user-owned profiles

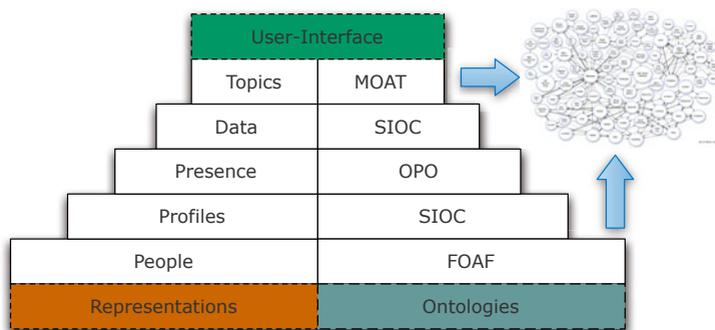
thanks to her of his URI. Then, for each microblog post, additional information about its author can be immediately obtained by dereferencing its author’s URI. Moreover, we also used OPO — Online Presence Ontology [20] — to describe users’ presence information, such as their geolocation aspects. Finally, we can benefit from these FOAF-based user profiles for authentication purpose, since SMOB hubs rely on FOAF-SSL [21] to enabling distributed authentication.

Regarding the modelling of microblogs and microblog posts, we relied on and extended SIOC — Semantically-Interlinked Online Communities [4]. Notably, we introduced two new classes to the SIOC Types module<sup>7</sup> to model these services and their data: (1) `sioc:Microblog` and (2) `sioc:MicroblogPost`. We also introduced two additional properties: (1) `sioc:follows`, to express following / follower notifications (the same property being used for both, benefiting of the RDF graph model), and (2) `sioc:addressed_to`, to represent whom a given post is intended for.

Finally, a further aspect concerns the enhancement of tagging practices, notably to bridge the gap between #tags as simple keywords and URIs identifying Semantic Web resources. We rely on MOAT — Meaning Of A Tag [16] — to do so, since it provide a model to represent the meaning of tags using Semantic Web resources, such as identifying that in a particular context, the tag `apple` is used to identify “apple, the fruit” (identified by `dbpedia:Apple`) but “Apple, the computer brand” (`dbpedia:Apple_Inc.`) in another one.

Combined together, these ontologies form a complete stack to represent the various elements involved in microblogging applications (Fig. 2). We did not want to provide a new unique and huge ontology, but rather defined this combination

<sup>7</sup> <http://rdfs.org/sioc/types#>, prefix `sioc`



**Fig. 2.** The SMOB ontologies stack

of lightweight ontologies [8] to fit with existing applications and provide a coherent framework to represent the different artifacts of paradigms appearing in the Social Web. Therefore, this stack is not specific to SMOB but can be referred to as a more global ontologies stack for the Social Semantic Web and can be reused in any applications relying on similar modeling requirements. Each post created with SMOB is then modelled using this stack and made available on the Web as RDFa-annotated content, in the author's hub. The following snippet of code (Fig. 3, prefixes omitted) shows an example of a microblog post represented via SMOB with this ontologies stack, here in Turtle.

```
<http://example.org/smob/post/20091101-1755> a sioc:MicroblogPost ;
  sioc:content "Drinking #coffee with @milstan in #Torino" ;
  sioc:has_creator <http://apassant.net/smob/owner> ;
  foaf:maker <http://apassant.net/alex> ;
  sioc:has_container <http://apassant.net/smob> ;
  sioc:addressed_to <http://ggg.milanstankovic.org/foaf.rdf#milstan> ;
  moat:taggedWith <http://dbpedia.org/resource/Coffee> .
<http://example.org/smob/presence/20091101-1755> a opo:OnlinePresence ;
  opo:customMessage <http://example.org/smob/post/20091101-1755> ;
  opo:currentLocation <http://sws.geonames.org/2964180/> .
```

**Fig. 3.** Representing microblog posts using the SMOB ontologies stack

### 3.2 Distributed Hubs and Synchronisation Protocols

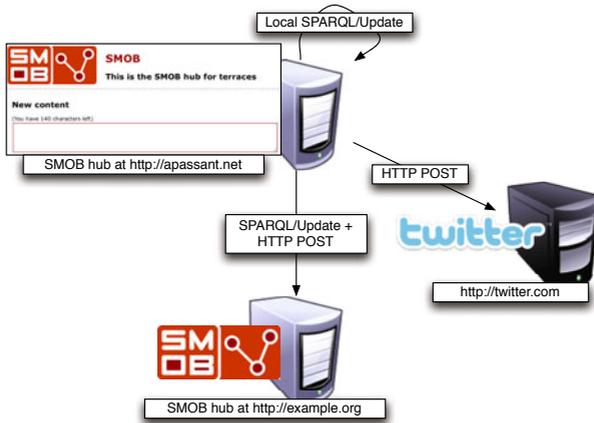
In order to fulfill our second requirement (R2: decentralised architecture and open data), we designed an architecture based on distributed microblogging hubs that communicate with each other to exchange microblog posts and notifications. That way, there is no centralised server but rather a set of hubs that contains microblog data and that can be easily replicated and extended. Hubs communicate with each other via HTTP thanks to SPARQL/Update (the Update part of SPARQL, currently being standardised in the W3C<sup>8</sup>). We rely in

<sup>8</sup> <http://www.w3.org/TR/sparql11-update/>

particular on a subset of SPARQL/Update, namely the **LOAD** clause, in order to publish items from one hub to another (Fig. 4). When creating a new microblog post on a SMOB hub, the workflow is the following:

- the post is immediately stored in the local RDF store, and published in an RDFa-enabled page at its own URI, e.g. <http://example.org/post/data>, using the aforementioned vocabularies;
- a SPARQL query identifies all the poster’s followers from the local store, as well as the URL of their hub;
- a SPARQL/Update query is sent (*via* HTTP POST) to each hub so that they aggregate the newly-created RDF data.

To avoid hijacking, only posts whose URI corresponds to the URI of a followee can be loaded in remote stores, so that only trusted data (*i.e.* generated by a followee) is aggregated. A SPARQL/Update pre-processor is used to provide this additional level of security on top of SPARQL endpoints. Future improvements may include FOAF-SSL or OAuth<sup>9</sup> to further address this issue.



**Fig. 4.** Communication between SMOB hubs using SPARQL/Update

Using this workflow, posts are immediately broadcasted to every follower as soon as the content is created. This push approach (contrary to a pull one where hubs would regularly fetch followees’ information) is similar to what is provided by the OStatus protocol<sup>10</sup>. However, while OStatus sends all information about the posts (using Atom feeds), our approach directly sends a SPARQL/Update **LOAD** query to each follower’s hub. Such a query only contains the URI of the post to be loaded, this URI being dereferenced when it is loaded in the follower’s RDF store. That way, we rely on “self-descriptive” posts as each URI

<sup>9</sup> <http://oauth.net/>

<sup>10</sup> <http://ostatus.org/>

identifying a microblog post can be dereferenced and delivers a full set of RDFa information about itself. This nicely illustrates the use of Linked Data principles to transmit rich status updates (since they can contain geolocation information, semantically-enhanced tagging, etc.) in the context of distributed microblogging.

In addition, each hub features a triggering approach which loads followers' FOAF profiles each time one of their posts is loaded into the store. Then, if a user edits his or her depiction, this will be forwarded to each hub following as soon as a new post is created, without any additional intervention, and without having to send this information in the post itself, thanks to the FOAF-based approach that we earlier described.

A similar approach is used with regards to the followers and followees subscriptions. A bookmarklet is provided to let anyone become a follower of another user when browsing that user's hub. The subscription is registered in the follower's hub by adding a "`<user-uri> sioc:follows <remote-user-uri>`" triple, while the same triple is included in the remote store. Thus, that both parties are instantaneously aware of this new relationship, which can be then used in the aforementioned publishing protocol when new updates are published.

### 3.3 Integrating Microblogging in the Linking Open Data Cloud

As we mentioned in our third requirement (R3: data-reuse and interlinking), our goal was to make microblog posts more discoverable by linking them to existing resources on the Web. In particular, our vision is to make microblogging updates linked to resources from the Linking Open Data cloud. While such interlinking is already provided by reusing existing FOAF profiles, there is a need to go further and link to relevant data to make sense of the content of microblog updates.

To achieve such goal, we extended the common `#tag` practice in order to turn these tags into machine-readable identifiers, *i.e.* URIs identifying Semantic Web resources. In particular, we aim at relying on URIs from the Linking Open Data cloud, since it provides identifier for various things, from generic topics in DBpedia (the RDF export of Wikipedia) to drug information information in LODD [12]. Our approach therefore focuses on extending the genuine `#tag` practice by turning tags into identifier to resources from the LOD cloud, enabling interlinking between tagged content and these resources. Then, `#tags` are not simple tags anymore, but provide links to uniquely identified and structured resources. In addition to the interlinking, this practice also solves the ambiguity and heterogeneity issues of tagging. Indeed, by linking a microblog post initially tagged with `#apple` to `dbpedia:Apple_Inc.`, one can identify that this post is about the computer brand (and not the fruit nor the record label), and could also benefit from the links existing from (and to) the related URI to identify that this is a message about computers (since `dbpedia:Apple_Inc.` and `dbpedia:Computer_hardware` are linked in DBpedia).

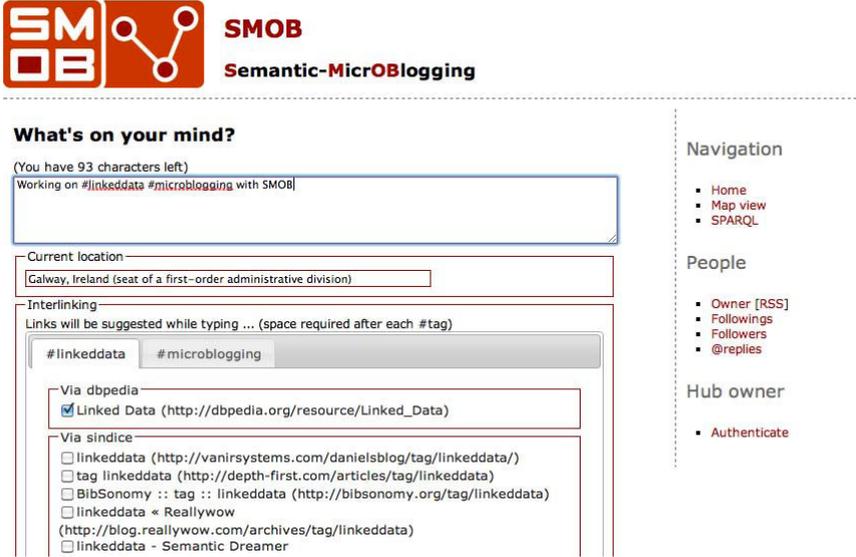


Fig. 5. The SMOB publishing interface and its interlinking components

## 4 SMOB in Use: Publishing and Discovering Semantically-Enhanced Microblog Posts

### 4.1 Publishing and Interlinking Microblog Posts

SMOB is available under the terms of the GNU/GPL license at <http://smob.me> and can be simply setup on any LAMP environment. Its user interface is somehow similar to existing microblogging applications, but when writing posts, a parser interprets the content entered by the user (using JQuery and a set of regular expressions) in order to identify if a know pattern has been used, such as **#tag**, **@user** or **L:loc** (generally used for geolocation purposes). In addition, each pattern is mapped to a set of wrappers that query existing services in real-time to suggest relevant URI(s) for the resource they may refer to.

Default wrappers for **#tag** include Sindice (the Semantic Web index<sup>11</sup>) or DBpedia, whilst also letting people write their own wrappers. The mappings are then modelled using MOAT and exposed in the microblog posts as RDFa, so that they can be used for querying as we shall see next. Regarding **@replies**, we have mainly relied on Twitter and on other SMOB hubs that the user interacts with so that we can link to a user's FOAF profile when responding to a message. Finally, for the **L:location** patterns, we use the GeoNames service<sup>12</sup> that provide URIs for more than six million geographic entities. Interestingly, not only cities or countries are referenced, but also various places such as hotels, which offers an interesting level of granularity for representing this information.

<sup>11</sup> <http://sindice.com>

<sup>12</sup> <http://geonames.org/export>

From the suggested URIs, users can then decide which one is the most relevant (Fig. 5), that choice being saved for further reuse. As soon as the post is saved, informations are stored in the local RDF store and can be browsed in the user interface using RDFa. Furthermore, the mappings are saved for next posts and can be shared between hubs so that one's mapping can be provided to followers to enhance the interlinking process. While this process of manual interlinking from `#tags` may sound complex, we recently demonstrated the usefulness of MOAT in a corporate context, showing how users could benefit from the system to improve information retrieval [16] and showed that users are willing to do the additional effort of assigning resources to their tags.

As new wrappers can be created, it may be useful to build these in enterprise contexts if people want to refer to their own knowledge bases. For example, a tag such as `#p:1453` could automatically be linked to the URI of the corresponding project, identified by its ID (1453). It therefore provides a use-case for enhancing microblogging in Enterprise 2.0 environments [15], following some of our previous work on Semantic Enterprise 2.0 [17].

SMOB can also be used as a Twitter client, relying on the Twitter API to do so. In that case, not only SMOB content is cross-posted to Twitter, but Twitter data is integrated in SMOB hubs and translated into RDF using the previous ontologies. Thus, Twitter data can be queried in a similar way to native SMOB data (using SPARQL) providing another way to make existing microblogging data enters the Linking Open Data cloud.

## 4.2 Geolocation Mash-Ups

In order to let users define geolocation information, SMOB enables a deep integration with GeoNames<sup>13</sup>. In addition to the webservice used to map the `L:loc` patterns, SMOB provides an autocompletion field so to users can define their current location. The auto-completion is based on the GeoNames webservice, and its JSON answers are interpreted on runtime to fill the location textbox. The main interest of GeoNames in our context is that each location has its own URI and has a description available in RDF. Thus, each time a new post featuring location information is created (or posted to a hub), the GeoNames URI corresponding to the current location (linked to the post with OPO, see previous example) is dereferenced and the data is integrated in the SMOB hub. We therefore benefit from any related information, such as the location of the feature, its inhabitants, its parent zone, etc. Consequently, posts can be geolocated in real-time, as seen in Fig. 6, and new features can be provided from the querying side, for instance identifying all content posted in a given country, while the only information available in the annotation refers to a city.

## 4.3 Data Discovery and Querying with SPARQL

In addition, SMOB hubs can be queried using SPARQL, either directly (via their local endpoint) or distributively, as each hub pings Sindice when new content

---

<sup>13</sup> <http://geonames.org>



Fig. 6. Real-time geolocation of microblog posts with SMOB and GeoNames resources

is created. Then, using a library such as the Semantic Web Client Library [9], which uses an approach to enable SPARQL query over the Web by traversing RDF graphs, one can discover hubs and microblog content that are distributed on the Web.

```

SELECT DISTINCT ?post ?author
WHERE {
    ?post a sioc:MicroblogPost ;
    sioc:has_creator ?author ;
    moat:taggedWith [ ?related dbpedia:Italy ] .
}
    
```

Fig. 7. Example of advanced SPARQL query retrieving SMOB data

As an example, the SPARQL query in Fig. 7 (prefixes omitted) identifies posts about Italy and their author, even if the tag #italy was not initially used in the post, but by benefiting of (1) the interlinking from the original post to the URI identifying a given location (e.g. dbpedia:Milan) and (2) the existing links between dbpedia:Italy and this location in DBpedia.

Additionally, it can provides means to subscribe to topic-based or geolocation-based feeds, without having to rely on various subscriptions feeds. Moreover, if the links in DBpedia are enriched, the feeds will be automatically updated with new resources.

## 5 Related Work

Since the first release of SMOB mid-2008, the first semantic microblogging platform developed, various related projects have emerged. Microblogging platforms generating semantic data (i.e. represented using RDF(S)/OWL) include

smesher<sup>14</sup> and StatusNet<sup>15</sup> (formerly Laconica). smesher is a semantic microblogging client with local storage, that integrates with Twitter and Identi.ca (another popular microblogging website, powered by StatusNet). StatusNet publishes both FOAF (describing people) and SIOC data (as SIOC-augmented RSS feeds for users and groups), and allows users to create friend connections across installations. It also uses the OpenMicroBlogging (OMB) protocol<sup>16</sup> for client-server communication (currently redesigned as OStatus). However, these applications do not provide interlinking with the Linked Data cloud, focusing only on representing the containers or structures using semantics, but not on linking its content to existing resources, hence leaving microblogging isolated from the ongoing LOD initiative. In addition, the semantics exposed in StatusNet are relatively light and do not take into account particular features of microblogging (such as modelling @user message recipients).

Moreover, in order to extend current microblogging systems, various syntax extensions have been proposed, including MicroTurtle<sup>17</sup>, microsyntax<sup>18</sup>, nanoformats<sup>19</sup>, Twitterformats<sup>20</sup> and TwitLogic [19]. These syntaxes however generally require specific ways of formatting microblog posts, that may be not be widely used and are restricted to particular niches of users.

In addition, one can also implement semantic capabilities on top of existing systems. SemanticTweet<sup>21</sup> provides exports of user profiles from Twitter using FOAF, and the Chisimba Twitterizer<sup>22</sup> provides microblog data using open formats, but the original content is still stored in a closed system. We also recently designed a system translating Twitter streams to RDF data (including some of the vocabularies used in the SMOB ontologies stack) in real-time<sup>23</sup> so that it can be used in Semantic Web applications using streamed SPARQL extensions such as C-SPARQL [1] However, many of these applications and syntaxes rely on the Twitter infrastructure, and do not offer the distributed and open architecture that SMOB provides.

## 6 Conclusion

In this paper, we detailed the architecture of SMOB, a framework for an open, distributed and semantic microblogging experience. In particular, we focused on the issues of existing microblogging systems and detailed how we designed and used (i) a set of ontologies combined together to represent metadata related to

---

<sup>14</sup> <http://smesher.org/>

<sup>15</sup> <http://status.net/>

<sup>16</sup> <http://openmicroblogging.org/>

<sup>17</sup> <http://buzzword.org.uk/2009/microturtle/>

<sup>18</sup> <http://www.microsyntax.org/>

<sup>19</sup> <http://microformats.org/wiki/microblogging-nanoformats>

<sup>20</sup> <http://twitterformats.org/>

<sup>21</sup> <http://semantictweet.com>

<sup>22</sup> <http://trac.uwc.ac.za/trac/chisimba/browser/modules/trunk/twitterizer>

<sup>23</sup> <http://realtimesemanticweb.org>

microblogs and microblog posts, associated with RDFa markup to model microblog data, (ii) a distributed architecture to make microblogging more open and let users claim their data, combined with a synchronisation protocol that can be reused in other distributed Social Semantic Web applications, and (iii) interlinking components to make microblogging — and more generally the Social Semantic Web — part of the ongoing Linking Open Data cloud and offer new querying and mash-up capabilities.

In future work, we may consider how to not only link to existing content, but to extract new information from microblog updates, following practices generally used in Semantic Wikis, and enabling in this case streamed statements creation and notification that could provide new possibilities in terms of real-time information monitoring.

## Acknowledgements

The work presented in this paper has been funded in part by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Líon 2). We would like to thank Uldis Bojārs and Tuukka Hastrup for their involvement in SIOC and in the first version of SMOB as well as Milan Stankovic and Philippe Laublet for their work on OPO.

## References

1. Barbieri, D.F., Braga, D., Ceri, S., Valle, E.D., Grossniklaus, M.: Continuous Queries and Real-time Analysis of Social Semantic Data with C-SPARQL. In: Proceedings of the Second Social Data on the Web Workshop (SDoW 2009). CEUR Workshop Proceedings, vol. 520, CEUR-WS.org (2009)
2. Berners-Lee, T.: Linked Data. Design Issues for the World Wide Web, World Wide Web Consortium (2006), <http://www.w3.org/DesignIssues/LinkedData.html>
3. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data — The Story So Far. International Journal on Semantic Web and Information Systems (IJSWIS) 5(3), 1–22 (2009)
4. Breslin, J.G., Harth, A., Bojārs, U., Decker, S.: Towards Semantically-Interlinked Online Communities. In: Gómez-Pérez, A., Euzenat, J. (eds.) ESWC 2005. LNCS, vol. 3532, pp. 500–514. Springer, Heidelberg (2005)
5. Breslin, J.G., Passant, A., Decker, S.: The Social Semantic Web. Springer, Heidelberg (2009)
6. Brickley, D., Miller, L.: FOAF Vocabulary Specification. Namespace Document, FOAF Project (September 2, 2004), <http://xmlns.com/foaf/0.1/>
7. Golder, S., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of Information Science 32(2), 198–208 (2006)
8. Gómez-Pérez, A., Corcho, O.: Ontology languages for the Semantic Web. IEEE Intelligent Systems 17(1), 54–60 (2002)
9. Hartig, O., Bizer, C., Freytag, J.-C.: Executing SPARQL queries over the web of linked data. In: Bernstein, A., Karger, D.R., Heath, T., Feigenbaum, L., Maynard, D., Motta, E., Thirunarayan, K. (eds.) ISWC 2009. LNCS, vol. 5823, pp. 293–309. Springer, Heidelberg (2009)

10. Huberman, B.A., Romero, D.M., Wu, F.: Social networks that matter: Twitter under the microscope. Computing Research Repository (CoRR), abs/0812.1045 (2008)
11. Java, A., Song, X., Finin, T., Tseng, B.: Why We Twitter: Understanding Microblogging Usage and Communities. In: Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007 (2007)
12. Jentzsch, A., Zhao, J., Hassanzadeh, O., Cheung, K.-H., Samwald, M., Andersson, B.: Linking Open Drug Data. In: Linking Open Data Triplification Challenge 2009 (2009)
13. Knorr-Cetina, K.D.: Knorr-Cetina. Sociality with objects: Social relations in post-social knowledge societies. *Theory, Culture and Society* 14(4), 1–30 (1997)
14. Mathes, A.: Folksonomies: Cooperative Classification and Communication Through Shared Metadata (2004)
15. McAfee, A.P.: Enterprise 2.0: The Dawn of Emergent Collaboration. *MIT Sloan Management Review* 47(3), 21–28 (2006)
16. Passant, A., Laublet, P., Breslin, J.G., Decker, S.: A URI is Worth a Thousand Tags: From Tagging to Linked Data with MOAT. *International Journal on Semantic Web and Information Systems (IJSWIS)* 5(3), 71–94 (2009)
17. Passant, A., Laublet, P., Breslin, J.G., Decker, S.: SemSLATES: Improving Enterprise 2.0 Information Systems using Semantic Web Technologies. In: The 5th International Conference on Collaborative Computing: Networking, Applications and Worksharing. IEEE Computer Society Press, Los Alamitos (2009)
18. Sheth, A.: Citizen Sensing, Social Signals, and Enriching Human Experience. *IEEE Internet Computing* 13(14), 80–85 (2009)
19. Shinavier, J.: Real-time SemanticWeb in = 140 chars. In: Proceedings of the WWW 2010 Workshop Linked Data on the Web (LDOW 2010), CEUR Workshop Proceedings. CEUR-WS.org (2010)
20. Stankovic, M.: Modeling Online Presence. In: Proceedings of the First Social Data on the Web Workshop. CEUR Workshop Proceedings, vol. 405. CEUR-WS.org (2008)
21. Story, H., Harbulot, B., Jacobi, I., Jones, M.: FOAF+TLS: RESTful Authentication for the Social Web. In: First International Workshop on Trust and Privacy on the Social and the Semantic Web (SPOT 2009) (June 2009)